Easing the Migration from Parallel to Serial Storage

White Paper
April 2004
Paul Griffith – Adaptec, Inc.

Introduction

Rapid data growth among companies of all sizes is sending IT departments scrambling for cost-effective ways to store more information under tightening budgets. A new class of data – reference information – is driving much of this growth. This infrequently accessed data such as email, presentations, documents, graphics and variety of images including CAD/CAM drawings, medical X-rays and bank checks is placing especially acute cost pressures on small and midsize businesses (SMBs), the market segment accounting for the bulk of reference data growth but the one typically with the smallest IT budgets. The result: SMBs, along with their enterprise counterparts, are seeking out cost-effective alternatives to enterprise-class disk drives for this exploding class of information. They also continue to explore ways to reduce the cost of storing high-demand database information and other bandwidth-intensive transactional data.

New storage technologies with serial architectures that provide higher performance and capacity scalability and enable denser designs are emerging to streamline IT operations and reduce total storage ownership costs. The biggest hurdle, however, to adopting new technology is the painful process of moving away from legacy systems that are not backward compatible with new ones. In particular, the challenge today for many organizations lies in moving mission-critical data from existing parallel-technology storage systems to new serial-architecture storage solutions with little or no downtime. Switching over to new systems in enterprise environments can be especially challenging because of the difficulty of scheduling downtime when many application servers need to be kept running round-the-clock.

The Move to the Serial Interface

For more than 20 years, data centers have relied on the parallel bus interface as the workhorse disk drive interconnect for both reference and transactional data. Parallel technologies – SCSI for enterprise environments and ATA for cost-sensitive desktop applications – remain in widespread use and continue to meet the performance requirements of many of today's enterprise solutions. However, significant technical challenges have made it economically impractical to use parallel technologies to meet increasing demand for more robust data integrity, higher system performance, greater storage flexibility and scalability, and smaller system designs.

Serial architectures have now emerged to deliver higher performance by allowing more bandwidth per device pathway than their parallel counterparts. Serial bus architectures also support a network of dedicated point-to-point device connections to provide full bandwidth to each storage device, unlike multi-drop parallel bus architectures, which share total bandwidth among devices. This dedicated serial connection also eliminates the single point of failure found in today's parallel environments, increasing reliability.

Serial storage technology, specifically Serial ATA and Serial Attached SCSI, address the architectural limitations of their parallel counterparts to deliver highly scalable performance. The technology draws

its name from the way it transmits signals – in a single stream, or serially, compared to multiple streams for parallel. The main advantage of serial technology is that while it moves data in a single stream, it wraps data bits into individual packets that are transferred up to 30 times faster than parallel technology data. In addition, serial technology's point-to-point architecture features dedicated connections that deliver full bandwidth to each device.

SATA and SAS Features

Serial ATA extends the parallel ATA technology roadmap by delivering disk interconnects speeds starting at 1.5 Gb/sec (150 MB/sec). Due to its lower cost-per-gigabit, Serial ATA will continue as the prevalent disk interface technology in desktop PCs, and sub-entry level servers and networked storage systems requiring low-cost storage.

Serial Attached SCSI, the successor technology to parallel SCSI, leverages proven SCSI functionality and features while expanding SCSI's proven performance, scalability and reliability for enterprise storage. Serial Attached SCSI offers many features not found in today's mainstream storage solutions such as drive addressability up to 16,000 devices per port, and reliable point-to-point serial connections at first-generation speeds of up to 3 Gb/sec. In addition, Serial Attached SCSI's small connector supports full dual-ported connections on 2.5-inch hard disk drives, a feature previously found only on larger 3.5-inch Fibre Channel disk drives. Dual-ported connections are essential for applications that require redundant drive spindles in a dense server form factor such as blade servers.

The Serial Attached SCSI interface also is compatible with lower cost-per-gigabit Serial ATA drives, giving system builders the flexibility to integrate either Serial Attached SCSI or Serial ATA devices and substantially reduce procurement, inventory and other costs associated with supporting two separate interfaces.

Challenges of Migrating Data

Migrating data to new serial-based systems can be complex and labor-intensive because of the myriad of application servers, operating systems, file systems, volume management, storage devices and networks at play in today's computing environments. The most formidable challenge IT departments face in migrating data is protecting data against loss and corruption and minimizing disruptions to ongoing business operations.

Since parallel-technology disk drives are incompatible with serial-technology interfaces, data can't be migrated by simply sharing storage across common interface controllers and copying the data. Instead, the new serial storage subsystem must be attached on a separate controller within the server or on a separate storage server altogether.

When it's not possible to establish a direct connection between existing and new storage, data can be migrated using network infrastructures such as LAN, WAN and SAN in conjunction with mirroring technologies designed to ensure continuous data availability.

Traditional Migrating Options

There are two traditional approaches to migrating data: stage data from the old storage device on either tape or another storage medium and restore it to the new device or using host-based software. These typically involve significant system downtime. With data migration, an IT manager will typically schedule an outage for a storage upgrade, though determining a downtime that will be the least disruptive can be very difficult.

Traditional migration – offloading data to a temporary storage device – involves manually taking a point-in-time copy of data and staging it for copying to the temporary device, installing the new storage and copying data from the temporary to the new device, a process that is often simple but can involve significant user downtime. To maintain data consistency, user access to stored data generally must be

interrupted throughout backup and restoration, a process that can't be used in environments requiring continuous data access

A less-intrusive alternative is to migrate data using host-based software that provides users with some access as data is being copied. The host software captures any I/O operation performed on the old storage system and replicates it on the new storage devices.

Utilizing the Network for Data Migration

A number of storage infrastructures are well entrenched today including Direct Access Storage (DAS), Network Attached Storage (NAS), and Storage Area Networks (SAN). In a DAS configuration, a parallel SCSI or ATA or other disk interface connects directly to the storage device. NAS storage devices are essentially file sharing engines attached to an existing Ethernet data network. A SAN is a networked storage infrastructure dedicated solely to storage. In a SAN, storage devices are attached to a network, traditionally a logically independent Fibre channel network, rather than directly to a server. In each of these networked environments, organizations can use emerging storage management tools such as storage virtualization, DAS spillover and DAS evacuation to migrate and expand storage in ways that are transparent to users, ensuring continuous data access.

Storage virtualization involves combining direct- and fabric-attached storage into a logical, or virtual, pool that the user sees as direct-attached storage to simplify allocation and management. The technology separates logical storage from physical storage by adding a vendor- and platform-neutral management layer between the servers and storage devices. Storage virtualization gives organizations an easy way to add new storage technology such as Serial Attached SCSI to an existing storage infrastructure, logically display the two distinct subsystems as one, and migrate data to the new storage capacity.

With DAS spillover, a DAS volume is extended to fabric-attached storage to create a virtual storage pool that, like virtualization, appears as direct-attached storage to the user. DAS spillover can be performed independent of the application, operating system and DAS vendor. It gives fabric-attached storage the same look and ease of manageability as DAS storage by using existing management tools. Chief among its benefits, DAS spillover does not disrupt the existing infrastructure, significantly reducing the management problems stemming from exhausting or expanding DAS storage. DAS spillover emulates DAS storage management, giving users complete control to migrate to fabric-attached storage.

IT managers also can simplify data migration by using Information Lifecycle Management (ILM) software to automate DAS spillover and offload reference data to fabric-attached storage. This frees up valuable direct-attached storage space and speeds access to transactional information in a process called cold data migration or adaptive storage migration. Depending on the sophistication of the ILM software, this technique allows either block- or file-level data in direct-attached storage to be allocated to fabric-attached storage based on access frequency and application need. Adaptive data migration can be executed in a variety of ways. The user, for example, can initiate the migration manually or establish a schedule for migrating data automatically at regular intervals. IT managers also can set the ILM software to migrate data automatically in real-time back and forth between the host and fabric and predict what data will be accessed next based on user-access patterns to optimize overall system performance.

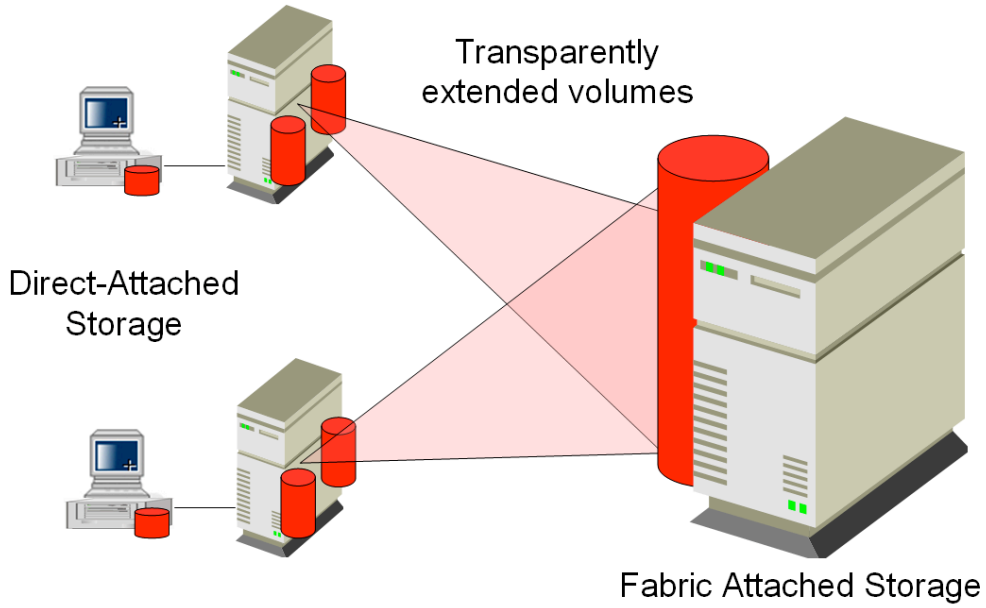

Figure 1: Using DAS spillover technology to seamlessly and transparently expand direct attached storage into fabric attached environments

DAS evacuation is the transparent and wholesale migration of DAS data to a fabric-attached storage device. The spillover region on the fabric-attached storage device must be expanded to accommodate the additional data. To the user, the operation is transparent and provides instant access to the new storage capacity. Evacuation decommissions DAS and moves the data to a fabric environment.

Conclusion

Today, most data centers use parallel storage technology like SCSI, but as requirements for increased performance, higher scalability needs and improved reliability grow, IT managers are looking to new serial technologies such as Serial Attached SCSI. IT managers needing a simple, cost-effective way to migrate from parallel-technology to serial-based systems with minimal disruption to business operations will find DAS Spillover, Adaptive Storage Migration, and DAS evacuation powerful tools for enabling this transition.

About the author

Paul Griffith is a Strategic Marketing Manager at Adaptec Inc. For more information, visit: http://www.scsita.org.